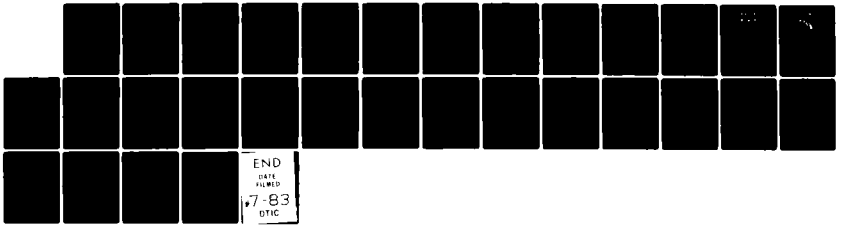
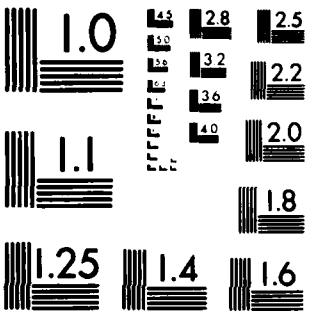
AD-A129 480 RECIPROCAL NEURAL PATHWAYS AND ASSOCIATIVE NETWORKS(U) 1/1
SALK INST SAN DIEGO CALIF F H CRICK ET AL. 15 DEC 82
AFOSR-TR-83-0542 AFOSR-82-0042

UNCLASSIFIED F/G 6/16 NL



END
DATE FILMED
7-83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963 A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

12

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 83-0542	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): *Reciprocal Neural Pathways and Associative Networks*	5. TYPE OF REPORT & PERIOD COVERED: 1 Nov 81 - 31 Oct 82 FINAL REPORT
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): F. H. C. CRICK and G. J. MITCHISON	8. CONTRACT OR GRANT NUMBER(s): AFOSR-82-0042
9. PERFORMING ORGANIZATION NAME AND ADDRESS: THE SALK INSTITUTE, PO BOX 85800, SAN DIEGO, CA 92138	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F, 2312/A1
11. CONTROLLING OFFICE NAME AND ADDRESS: USAF OFFICE OF SCIENTIFIC RESEARCH/NL, Bolling AFB, DC 20332	12. REPORT DATE: 12/15/82
	13. NUMBER OF PAGES: 23
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

DTIC ELECTE S JUN 20 1983 D B

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Dendritic Spines, Cerebral Cortex, Visual Cortex, Memory, Rapid Eye Movement Sleep, Neurophysiology

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This ~~final~~ report of ~~approximately one year's work~~ covers four separate but related topics. The first concerns the suggestion that dendritic spines may "twitch" -- that is, change shape rapidly. This has already been published and appears as Appendix A. The second concerns the patterns of long-range connections in the visual cortex. This also has been published and is reproduced as Appendix B. The third topic concerns the problem of memory storage at higher levels in the cortex. This is in a preliminary

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

83 06 20 040

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A129480

DTIC FILE COPY

stage and only a very broad account is given here. The fourth topic is the most speculative and concerns the function of Rapid Eye Movement Sleep and the nature of dreams. A draft paper on this is reproduced as Appendix C.

The body of the report gives brief accounts of all these topics and shows how they are linked together.

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A

FINAL TECHNICAL REPORT
Air Force Grant AFOSR-82-0042

RECIPROCAL NEURAL PATHWAYS AND ASSOCIATIVE NETWORKS

F. H. Crick and G. J. Mitchison

The Salk Institute

10010 N. Torrey Pines Road

LaJolla, California 92037

Controlling Office: USAF Office of Scientific Research/NL
Bolling AFB, DC 20332

Approved for public release;
distribution unlimited.

NOTI
Thi
ap
Di
MA
Chief, ivision

The general nature of the cortex

The anatomy of the cortex shows a clear general pattern of connectivity: Cells make rich connections to their neighbors over a distance of about 1 mm, and there are also larger scale connections which run between different cortical 'areas', these being anatomically or functionally defined regions which seem to specialize in some processing task. There seems to be a broad overall flow of processing. For instance, about a dozen visual areas have been identified in the macaque, and these can be placed in a rough flow diagram, beginning with the striate cortex. As one might expect, given the limitations of local connections, receptive fields of cells in the early stages of visual processing are small (in the retina, lateral geniculate and striate cortex), and those in the later stages are much larger. This suggests that a general process of abstraction might take place in vision and the other senses, whereby simple and spatially localized features are detected, then more elaborate and less spatially localized features generated from them. In this way, a representation of an object could be constructed which would be independent of its exact position in space.

This philosophy was adopted by theoreticians who worked on 'perceptron' models of nerve networks. Their hope was that they could make computational models based upon idealized nerve cells (perceptrons) which could solve problems like position invariance. As it happened, they made rather little progress, and we can now see that one major difficulty lay in the fact that creating a succession of ever higher order features leads to profound combinatorial problems. Some other mechanism is needed which can say which features belong together. One important suggestion, which comes from psychophysical experiments by Treisman (1980) and Julesz (1981), is that there is a local attentional focus, a sort of internal moving 'spotlight', which sequentially surveys a scene and picks out features which lie close together. We believe this to be a very important hypothesis, whose implications have yet to be fully understood.

Another suggestion was made by von der Malsberg, who proposed that features could be bound together by temporal correlations in the firing of cells which represent them. To make such a scheme work, he argued, synapses between cells should be activated on a very short time scale. So, if one cell made a synapse onto a second, and if their activity was correlated, then the synapse should become effective and increase the correlation between the two cells. This should happen over a time scale of a few tens of milliseconds, he suggested. Such very rapid activation of a synapse would have other uses. For instance, it could serve as a short term memory store, by marking the connections between a subset of cells which had been active together. But what kind of mechanism could produce the necessary rapid changes?

Dendritic spines and very rapid changes of synaptic weights

Many cortical neurons, including the pyramidal cells which are such a characteristic feature of the cortex, are richly supplied with dendritic spines. These are short, narrow processes, about 2 microns long, and between .2 and 1.0 microns in diameter, which project from dendrites. They generally expand slightly at their tip into a bulb, onto which synapses, usually excitatory, are made from neighboring axons. That part of the spine which is nearest the dendrite, the 'neck' of the spine, can vary in width from .2 microns to about 1 micron. Rall suggested that variation in the width of this neck could alter the efficacy of a synapse on the bulb, essentially by increasing the resistance of the pathway from synapse to cell body, thus changing the amount of 'choking back' of the brief ionic current induced by the activation of the synapse.

Returning now to the question of making rapid changes in synaptic weight, F.H.C.C. suggested that these could be generated if the width of the neck of a spine could be rapidly altered. He proposed that the diameter of the neck is controlled in part by contractile proteins such as actin. This paper is reproduced as Appendix A. Since its publication several groups (Fifkova and Delay 1982; Katsumara et al., 1982) have shown that there is indeed actin in spines, possibly a special type of actin unlike that found at the synapse itself or in the neighboring dendrite (Caceres, et al). This is an encouraging finding, though of course it does not yet prove the hypothesis.

Oriented axons in the striate cortex

We turn next from general properties of the cortex to the specific features of early visual processing. It is now over 20 years since Hubel and Wiesel discovered that most cells in the striate cortex responded best to oriented edges or bars, and that there are two main types of cell - the simple and complex. It is not yet known what neuronal machinery gives cells their characteristic receptive field properties. However, it is clear that a highly organized topographical arrangement of connexions must be present, since it is known that cells of particular orientations are arranged in stripes and other patterns on the cortex. We were therefore very intriged to read a paper by Rockland and Lund, in which they announced that a pattern of stripes could be produced by simply injecting horseradish peroxidase (HRP) into the striate cortex of the tree shrew. This enzyme is known to be carried along axons of neurons, so that the location of the stain can reveal organization in the arrangement of the axons.

They had suggested that the tree shrew striate cortex contained two systems of connections, one having a long range

and the other only a short range. Their idea was that these systems formed alternating stripes, and that the HRP pattern was generated primarily by the axons of the long range system, running from the injection site to a number of stripes nearby.

Now, in the tree shrew striate cortex, cells of like orientation are arranged in roughly parallel stripes with a spacing of about 1mm. It is thought that stripes of different orientation intercalate, so that orientation changes smoothly and systematically in the direction perpendicular to the stripes. In dimensions and orientation, these stripes roughly resemble those of the HRP pattern. This raised the question of whether there was some relationship between the two systems. One possibility would be that cells of like orientation are joined together. Then, if the HRP were taken up by a cell of one particular orientation only, and if the label propagated to others in the same and adjacent orientation stripes, Rockland & Lund's pattern could be explained. However, this cannot be a correct explanation, since the injection site is quite large, and almost certainly includes cells of all possible orientations. This would lead to a labelling of all the orientation stripes, which would produce a uniform non-stripey pattern.

Our proposal, reproduced here as Appendix B, was that cells of like orientation were indeed connected to each other, but only in a particular orientation on the cortex, corresponding to the cell's preferred orientation when translated into visual space. This has a very natural interpretation, for it allows long, oriented receptive fields to be constructed out of short, oriented fields. It could also be used to generate complex receptive fields from simple ones by joining cells side to side rather than end to end.

Will such a pattern of connectivity generate stripes like those seen by Rockland and Lund? The Figures in our paper (Appendix B) show that they will. The HRP stripes will not be identical with the orientation stripes, but will cut across them at a small angle, which depends upon the position on the cortex with respect to the HRP injection site. It is interesting that recent evidence suggests that the two stripe systems are similar, but do not precisely coincide (Rockland et al 1982), which is what we should predict.

Some recent evidence from Gilbert and Wiesel (1982) has added substance to our view. They filled individual cells with HRP, then reconstructed the cells in 3 dimensions. When viewed from above, many of the cells had highly elongated axons, and showed moreover local clusterings of axon ramifications, as though rich connections were being made in patches. This again would fit with our interpretation of the Rockland & Lund pattern, as we should expect to see long axons making patchy connections.

In summary, there are two broad hypotheses. The first, due to Rockland and Lund, postulates that regions with long connections are interspersed between regions with only short

connections. The second, which we proposed, is that all regions have long connections, but only in particular directions. More experimental evidence will be needed to decide between these two hypotheses. It is possible that both may be correct to some extent.

Associative networks in the cortex

We have a general interest in the kind of connection patterns which might underly an associative memory, since this is undoubtedly another aspect of cortical functioning. In a certain sense, there is little doubt that the striate cortex 'learns', as shown by experiments on monocular deprivation and on raising of kittens in special environments. However, the learning which occurs only takes the form of variations within a prescribed pattern. The set of ocular dominance stripes for one eye expands at the expense of the others, for instance. Perhaps all cortical learning is like this, but it seems reasonable to suppose that at later stages, when information is more highly processed and 'symbolic', there is a less rigid prescription of an underlying order, and entirely novel patterns of connections can be made.

Let us therefore imagine a region of cortex whose inputs take the form of highly processed polymodal sensory information. An example of such a region might be the inferotemporal cortex, or the hippocampus. Suppose a subset S of cells in such a region are active, and we wish to record this activity in some way so that when a fraction of the active cells is subsequently re-activated, then the whole of S becomes active. An obvious way to do this is to make reciprocal synaptic connections between all the cells in S. It can be shown that such an "auto-associative" memory can store many inputs, with a total information content of about $\log_e N$, where N is the number of modifiable synapses in the memory system (Longuet-Higgins et al., 1970; Palm 1980). Can one make an estimate of the number of such inputs which might be stored in the cortex, and say how they should be distributed?

Using estimates for the number of synapses made on the dendrites of cortical pyramidal cells, and on the area of cortex over which they are likely to make rich connections, G.J.M. estimated that about 10^5 subsets of cells could be stored in a region of about $1mm^2$ of cortex, and that there was no gain in the <u>number</u> of subsets which could be stored by using larger areas of cortex. Larger areas would give larger and more information-rich subsets, but would not allow more than 10^5 subsets to be stored. As this number seems grossly low for human memory capacity, he suggested that there is a control mechanism which allows the cortical system to store memories in particular locations in isolation from other storage sites, and to retrieve the information in a similar manner. This would lead to the view that some regions of the cortex are subdivided into many "patches", within each of which memories are stored

without interference from neighboring patches.

This work is still in a very preliminary stage, so that it would be inappropriate to give a more detailed account here. When it is further developed and reaches publication, acknowledgement will be given to the support given by the Air Force.

Dream sleep

We now come to what is perhaps the most speculative of our ideas. While investigating the properties of associative memories, it became clear to us that such a system could malfunction if the subsets it was given to 'memorize' tended to overlap, or to be clustered over a small fraction of the total ensemble of cells. One kind of solution to the problem was to reprogram, or suitably transform, the inputs so that the clustering was reduced. However, it is not easy to see how this could be done effectively. F.H.C.C. proposed a different kind of mechanism, according to which the system is presented with a *pulse of noise* and allowed to respond. Under these conditions, the clustered cells will tend to respond particularly strongly, and we can regard their behaviour as a kind of 'parasitic mode' of the system. If the synaptic connnections between the cells which are excited by such noise are subsequently weakened, then this mode will be eliminated. As memories are distributed, and involve many connections with many different cells, this weakening of synapses will not destroy memories, and may in fact improve them, by eliminating 'parasitic' responses.

We propose that such a process could take place during dream sleep. It is thought that the marked EEG activity during REM sleep is generated by groups of cells in the brainstem, which deliver pulses of activity over large areas of cortex. This we would regard as the source of our pulse of noise for testing the system. We speculate that the weakening of synapses occurs by a process of reverse learning. This might be like the usual Hebb rule for synaptic learning, but with the sign reversed. Thus, if a synapse was active and the postsynaptic cell also active, then the synapse would be weakened, rather than strengthened as the Hebb rule requires.

Our draft paper on this topic is attached as Appendix C. It has recently been submitted for publication.

References

Caceres, A., Payne, M.R., Ruider, L.I., and Steward, O. 1983 (ms submitted for publication).

Fifkova, E. and Delay, R.J. 1982. Cytoplasmic actin in neuronal processed as a possible mediator of synaptic plasticity. J Cell Biol 95, 345-350.

Gilbert, C. D. and Wiesel, T. N. 1982. Clustered intracortical connections in cat visual cortex. Society for Neuroscience, Abstracts 204.6.

Julesz, B. 1981. Textons, the elements of texture perception and their interactions. Nature 290, 91-97.

Katsumaru, H., Murakami, F., and Tsukahar, M. 1982. Actin filaments in dendritic spines of red nucleus neurons demonstrated by immunoferritin localization and heavy meromysin binding note. Biomed Res 3(3), 337-340.

Longuet-Higgins, H.C., Willshaw, D.J. & Buneman, O.P. 1970 Theories of associative recall. Q. Rev. Biophys. 3, 223-244.

Palm, G. 1980 On associative memory. Biol. Cyb. 35, 1-13.

Rockland, K.S., Lund, J.S. & Humphrey, A.L. 1982. Anatomical banding of intrinsic connections in striate cortex of tree shrews (Tupaia glis). J. Comp. Neurol. 209, 41-58.

Treisman, A. M. 1980. A feature-integration theory of attention. Cognitive Psychology 12, 97-136.

Reprinted from Trends in NeuroSciences – *February, 1982* Vol. 5, No. 2, pp. 44–46

Do dendritic spines twitch?

Francis Crick

In the cerebral cortex the majority of synapses are located not on the somata of the neurons, nor on the shafts of their dendrites, but on dendritic spines. I wish to put forward a novel hypothesis: that there are contractile proteins associated with each spine, probably in the cytoplasm, which allow the spine to change its shape rapidly during neuronal activity. On this picture a neuron possessing spines, such as a pyramidal cell or a spiny stellate cell, is not a static object, since many of its spines are changing their shapes in response to synaptic inputs.

Dendritic spines were first observed in the last century by Cajal, under high-power light-microscopy, using neural tissue stained by the Golgi method (see Fig. 1). On this scale a spine is small, its volume being comparable to a fraction of that of a bacterium such as *Escherichia coli.* This comparison shows that, on the molecular scale, spines are quite large objects. Using electron microscopy[7,11,12] its structure can be seen in much more detail. Spines come in many shapes and sizes, as can be seen in Fig. 2. The end remote from the dendritic shaft is often rather bulbous, being connected to the shaft of the dendrite by a somewhat narrower neck. Most important, and unknown to Cajal, the bulb always has on its surface a synapse (occasionally more than one). Such synapses can usually, but not always, be classified morphologically as Type 1 (Ref. 4) and are therefore believed to be excitatory. (Spines have a small number of Type 2 synapses, but they never occur on a spine unless there is a Type 1 synapse there as well.) An electron microscopic picture of a section of a spine is shown in Fig. 3.

It has been reported by Jones and Powell[7] that the cytoplasm of a spine looks a little different in the electron microscope from that in the adjacent dendritic shaft, the former being somewhat denser and also flocculent in appearance. Peters, Paley and Webster[12] state that, 'In the electron microscope it is found that the spines in the cerebral cortex are filled with a fluffy material which seems to consist of fine and indistinct filaments, and this material is so characteristic that isolated profiles of the bulbs of spines can be readily recognized in the neuropil.' A peculiar structure known as a spine apparatus is often found in or near the neck of each spine. Spines rarely contain either mitochondria, microtubles, neurofilaments of ribosomes, all of which can usually be seen in some parts of the dendritic shaft itself.

Neurons in the fully developed cerebral cortex can be roughly divided into two broad classes: those with many spines and those with few or none. The majority of cells with spines are either pyramidal cells (of which there are many sub-types) or stellate cells. Together these two classes of spiny cells account for perhaps 90% of the neurons of the cortex. A typical stellate cell may have 300–500 spines on its dendrites, a typical pyramidal cell perhaps ten times as many, a very large pyramidal cell perhaps ten times more than that. In fact the great majority of Type I (excitatory?) synapses in the cerebral cortex are on spines. Thus, spines are not an occasional structural feature of cortical neurons. Their presence clearly demands some explanation, especially as in many other parts of the brain they are either absent or at least much less common.

What are spines for?

A number of authors have suggested one function or another for dendritic spines. Some of these have been reviewed by Diamond, Gray, and Yasarsil[2]. For example, it has been claimed that the neck of the spine isolates the synapse from the electrical effects of neighboring synapses. Recent calculations (Koch and Poggio, unpublished), suggest that this effect is weak or absent. More recently, Swindale[17] has proposed that spines are necessary if axons are to contact dendrites efficiently, especially when the density of such contacts is high. A similar suggestion was made earlier by Peters and Kaiserman-Abramof[11]. This idea, as Swindale points out, does not preclude other functions for spines, nor does it, by itself, explain why the cytoplasm of a spine should be different from that of the dendritic shaft to which it is attached

Perhaps the most interesting suggestion springs from a remark by Chang[1] in 1952, that the electrical resistance of the neck of the spine could reduce the 'weight' of the synapse: that is, the electrical effect which a presynaptic action of the synapse has on the impulse initiation site of the cell (the initial segment of the axon). This is not a straightforward attenuation effect but is due to the fact that synaptic activation involves a transient change of conductance at the synapse[15,16]. The extra resistance of the spine neck chokes back the resultant flow of ions during synaptic activation. For these general reasons, Rall and Rinzel[14] suggested in 1971 that memory might in part be stored in the shape of the dendritic spines. Subsequently, Rall[13] pointed out that the input impedance of the spine tends to match that part of the dendrite to which it is attached, since spines with long narrow necks are found more often on the distal parts of the dendrites.

More sophisticated calculations by Koch and Poggio (unpublished) have confirmed Rall and Rinzel's theoretical conclusions[14–16]. These calculations assume that the dendritic membrane is passive. The electrical parameters put into the calculations are known only approximately but plausible values for a change of shape could alter the weight of the synapse by a factor of

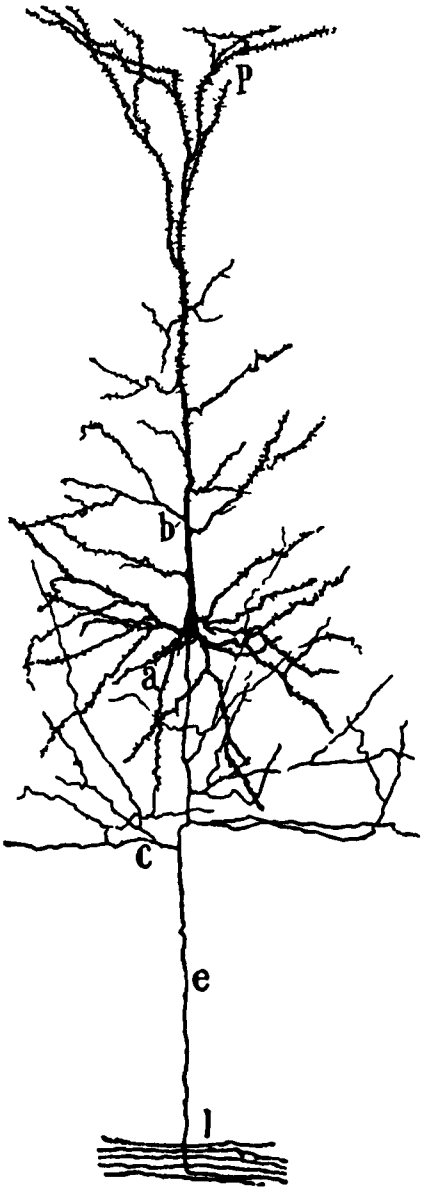


Fig. 1. *A drawing from Cajal showing a typical pyramidal cell stained by the Golgi method. The spines are the little twigs on the dendrites.*

© Elsevier Biomedical Press 1982 0378 - 5912/82/0000 - 0000/$02.75

two or three. Thus, the effect may be quite appreciable.

There is one possible difficulty with this idea. A spine does not appear to have a very rigid structure. Moreover, spines come in many different shapes and sizes. Yet long-term memory, in a human being, may last for several decades. This is remarkable in the face of the general metabolic turnover of all known macromolecules (DNA is probably an exception) and might suggest that long-term memory resides in a more rigid structure, with internal subunits arranged in a regular lattice, such as the synaptic junction itself, so that old molecules could be replaced by new ones without perturbing the general structure too much. However, since the molecular basis underlying the shape of a spine is not known, this objection should not be given too much weight.

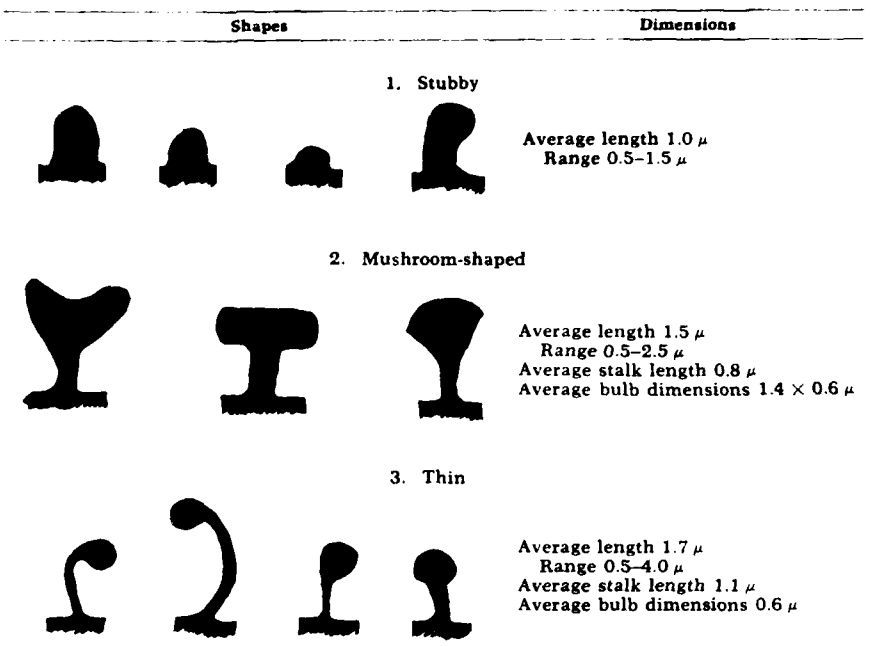


Fig. 2. *Shapes and lengths of dendritic spines.* (From Ref. 11.)

Theories of memory

I use memory here in the loosest sense, meaning any alteration in the nervous system which changes its subsequent performance. I do not address the problem of recall.

The nervous system of higher animals consists of many highly interconnected nets of nerve cells, interconnected with each other. It is generally believed that the behavior of such nets depends crucially on these elaborate patterns of interconnection and that long-term memory is stored, in a distributed and redundant way, as the weights of the many synapses. Using this view, a new memory is recorded as a small alteration to the weights of many synapses.

Various rules have been proposed for the manner in which synaptic weights are altered. A very popular rule, attributed to Hebb[6], is that a synapse is strengthened (that is, the synaptic weight is increased) if the presynaptic activation of the synapse is approximately coincident with the firing of the cell post-synaptic to the synapse. A number of more elaborate rules along these lines have been suggested. They all have in common that the weight of each synapse is either increased or decreased depending on the correlation between the presynaptic activation and the post-synaptic activity. There is rather general agreement that some sort of rule of this type (or the equivalent in local circuitry) would provide a great advantage in the efficient functioning of higher nervous systems. The typical times over which effective modifications are likely to occur have usually been seconds, minutes, or even longer, depending on the phenomenon being studied.

Ultra-short memory

Recently a new idea has been suggested. I first learned of it from von der Malsburg[18], sitting round a campfire in Aspen, Colorado, but a similar suggestion (though from a somewhat different viewpoint) had been made in 1975 by Little and Shaw[10], i.e. that it would be a considerable theoretical advantage (for example, in visual perception) if the weights of synapses could be changed, temporarily, by a large amount in much shorter times (of the order of, say, 10 ms) according to some Hebbian or Hebbian-like rule.

It is widely believed that very short-term memory depends in some way on reverberating electrical circuits or resonance but it has usually been assumed that such memory is expressed mainly by the firing patterns of complex sets of neurons and involves little or no change in synaptic weights, which are only altered appreciably when long-term memory is laid down. The novelty of these recent ideas about ultra-short memory is that they postulate that during neural activity there is a rapid but temporary change in the weights of the synapses involved, so that the required resonances can be built up more efficiently.

This suggestion poses an immediate difficulty. It has never been easy to see how an activated synapse could 'know' that the (post-synaptic) cell had fired. How is this information conveyed from the initial segment of the axon back to the synapse, a distance which can easily amount to several hundred microns? Given enough time the problem seems less troublesome but if it is to be done in 10 ms the problem becomes acute, since even small molecules cannot diffuse very far in such a time. For example, for a small molecule with a diffusion constant of 5×10^{-6} cm^2 s^{-1} the typical diffusion distance comes to 1 μm in 1 ms, or 10 μm in 100 ms. One possible solution might be that each time the cell fires an action potential is propagated from the soma towards the tips of the dendrites; there is, as yet, no evidence for this in cortical pyramidal cells.

The problem would be much easier if the unit of rapid learning were the dendrite rather than the cell. Recent calculations by Koch, Poggio and Torre[8] have suggested that, very loosely speaking, each dendritic branch may act to some extent as a separate logical unit in handling the incoming information in the synaptic activity. (The exact formulation is very much more complicated than this simple description.) More precisely, the rule for fast temporary modification of the weight of a given synapse might depend on the correlation between the activation of the synapses near to it on the dendrite. For example, it might depend on the potential in the dendritic shaft near the synapse at the time when the latter was activated presynaptically, since this potential would reflect the recent activity of neighboring synapses.

In von der Malsberg's formulation[18] the synaptic weight is the product of two terms, the first of which is unaltered in a very short time and represents the store of long-term memory, while the second can increase or decrease very quickly within defined limits. This rapid increase or decrease is controlled by some Hebbian or Hebbian-like rule. In the absence of all activity this latter term relaxes to some intermediate value.

Can dendritic spines change their shape?

It is immediately obvious, if this chain of ideas has any validity, that the neck of the dendritic spine is ideally suited to perform this function. By altering its shape it can contribute a suitable term to the synaptic weight. The contraction of an elongated gel is likely to make it shorter and fatter and thus decrease the electrical resistance of the neck. The position of the neck is such that it could, by one mechanism or another, be influenced both by its own synapse(s) and by any very recent activity of neighboring synapses. Because a spine has an unusual cytoplasm it is not outrageous to suggest that it might contain a fair amount of contractile proteins such as actin or myosin. Both the time needed for a muscular gel of this type to contract and the time for ions or small organic molecules to diffuse over the small distances involved are short enough for the proposed mechanism to work, as is the time for the contractile gel to relax when activity ceases.

It would thus seem very worthwhile to look with modern immunocytochemical techniques to see if the spine contains significant amounts of contractile proteins such as actin or myosin. It would also be sensible to check for the distribution of tubulin and the components of neurofilaments. Since antibodies are now available to all these proteins, such experiments should not be too difficult to do. Myosin is a rather variable protein, both between species and in different parts of the body within a species, but actin is relatively invariant. If there is indeed an appreciable amount of actin in dendritic spines it is a fair guess that it will react to at least some antibodies to other forms of actin. It would also be sensible to see if there are contractile proteins just outside the neck of the spine. For example, they might be arranged circumferentially so that, on contraction, the neck of the spine was constricted.

How compelling is this suggestion?

Two groups of authors[3,9] have presented experimental evidence that spines in the hippocampus can change as a result of the kind of massive stimulation which produces postactivation potentiation. Fifkova and van Harreveld[3] have suggested that these changes of shape are due to swelling produced by the action of glutamate (see Ref. 3 for earlier references) but the time scales involved are much longer (at least several minutes) than those being considered here. On the theoretical side it must be admitted that the advantages of ultra-fast synaptic modification within a nerve net are not yet completely clear. The idea is clearly attractive, as is the suggestion that the unit of modification is a stretch of dendrite,

Fig. 3. *An electromicrograph by Jones and Powell*[7]*, of the cat somatosensory cortex, showing a dendrite, marked d, with a spine marked s. The arrowhead shows the synapse.*

rather than the whole nerve cell, since both ideas would permit neural nets to behave in more complex ways. So far no compelling theoretical arguments have been produced, although some of them are suggestive. von der Malsburg has claimed that such a process is necessary in order to select one topological grouping of neurons from other possible ones within the same part of a neural net, but it could be argued either that such a process is not needed or, alternatively, that it is more likely to be done by special local circuitry, involving inhibitory neurons. Little and Shaw[10] feel that a rapid modification mechanism would greatly assist the type of short-term memory which depends on rehearsal for its maintenance but again their argument is only suggestive. Thus, it cannot yet be said on theoretical grounds that it is highly likely that a synapse will change its weight very rapidly. It is the extensive occurrence of dendritic spines in the cerebral cortex, in exactly the right position, which make the idea sufficiently plausible to be put to a preliminary experimental test.

Suppose actin was discovered in spines?

There is some evidence for actin being associated with post-synaptic structures at the neuromuscular junction[5] but none, as far as I know, for its presence in the cytoplasm of a dendritic spine. If it can be shown that actin is present in high concentrations in spines but absent, or in a rather low concentration, in the adjacent dendritic shaft, it would not only make us look at spine bearing neurons in a new light but would also open up many new fields for experimentation. Do dendritic spines twitch rapidly during normal neural activity? If so, what are the rules governing the change of shape of the spine and, in particular, the neck of the spine? Are they indeed of Hebbian-like nature or do they have some simpler character? (The change of shape might depend only on the presynaptic activity, for example.) How are these rules implemented in molecular terms? What is the spine apparatus for? How much energy is used up by the constant twitching? And so on.

All these speculations aside, the fact that the rather dense cytoplasm of spines seems to consist of fine, indistinct filaments surely makes it worthwhile to try to discover what this material is made of. A first step in that direction would be to check whether it contains known fibrillar proteins, of which actin is a good example.

Acknowledgements

I have benefited greatly from comments made by Max Cowan, Floyd Bloom, Graeme Mitchison, Bob Livingston and Tommy Poggio.

Reading list

1 Chang, H. T. (1952) *Cold Spring Harbor Symp. Quant. Biol.* 17, 189–202
2 Diamond, J., Gray, E. G. and Yasargil, F. M. (1970) in *Excitatory Synaptic Mechanisms* (Anderson, P. and Jansen, J. K. S., eds), pp. 213–222, Oslo Universitetsforlag
3 Fifkova, E. and van Harreveld, A. (1977) *J. Neurocytol.* 6, 211–230
4 Gray, E. G. (1959) *J. Anat.* 93, 420–423
5 Hall, Z. W., Lubit, B. W. and Schwartz, J. H. (1981) *J. Cell Biol.* 90, 789–792
6 Hebb, D. O. (1949) *Organization of Behaviour*, John Wiley and Sons, New York and London
7 Jones, E. G. and Powell, T. P. S. (1969) *J. Cell Sci.* 5, 509–529
8 Koch, C., Poggio, T. and Torre, V. *Phil. Trans. Roy. Soc. London* (in press)
9 Lee, K. S., Schottler, F., Oliver, M. and Lynch, G. (1980) *J. Neurophysiol.* 44, 247–258
10 Little, W. A. and Shaw, G. L. (1975) *Behav. Biol.* 14, 115–133
11 Peters, A. and Kaiserman-Abramof, I. R. (1970) *J. Anat.* 127, 321–356
12 Peters, A., Palay, S. and Webster, H. de F. (1976) *The Fine Structure of the Nervous System*, W. B. Saunders Company, Philadelphia, PA
13 Rall, W. (1978) in *Studies in Neurophysiology* (Porter, ed.), pp. 203–209, Cambridge University Press
14 Rall, W. and Rinzel, J. (1971) Cited in reference 15
15 Rall, W. and Rinzel, J. (1973) *Biophys. J.* 13, 648–688
16 Rinzel, J. and Rall, W. (1974) *Biophys. J.* 14, 759–790
17 Swindale, N. V. (1981) *Trends NeuroSci.* 4, 240–241
18 von der Malsburg, (1981) *Internal Report 81-2*, Department of Neurobiology, Max-Planck-Institut for Biophysical Chemistry, Goettingen

Francis Crick is at the Salk Institute, P.O. Box 85800, San Diego, CA 92138, U.S.A.

Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 3661–3665, June 1982
Neurobiology

Long axons within the striate cortex: Their distribution, orientation, and patterns of connection

(horseradish peroxidase/stripes/orientation columns/tree shrew/cat)

GRAEME MITCHISON AND FRANCIS CRICK

The Salk Institute, Post Office Box 85800, San Diego, California 92138

Contributed by Francis Crick, March 8, 1982

ABSTRACT **Rockland and Lund [Rockland, K. S. & Lund, J. S. (1982) *Science* 215, 1532–1534] have recently observed that an injection of horseradish peroxidase into the striate cortex of the tree shrew produces a patchy distribution of label adjacent to the injection site. They proposed that this pattern might be due to populations of neurons with long-range cortico-cortical connections that are interspersed with populations having no such connections. We suggest here an alternative explanation. We can account for the pattern by supposing that the label is carried by a system of oriented axons. We suppose that these axons link cells with similar orientation preferences and make their connections within a narrow strip of cortex whose direction is related to the orientation of the cells in question. We suggest that such connections could be involved in generating complex receptive fields from simple ones. Other possibilities are that they are used to generate very elongated receptive fields, inhibitory flanks, or end-stopping. We suggest a number of experimental tests of these ideas.**

In a recent paper, Rockland and Lund (1) made the remarkable observation that a local injection of horseradish peroxidase (HRP) in the primary visual cortex (also called the striate cortex or area 17) of the tree shrew gives rise to labeling not only in the immediate neighborhood of the injection but also in small patches all around the injection site. Their tentative interpretation was that there may be two intercalated systems of interconnections in area 17, one with long-range connections that produced the patches of HRP labeling, and a second that lacks them. We show here that their pattern could be explained in a different way, by supposing that neurons which have similar orientation preferences are joined together by a system of oriented axons. We shall see that this hypothesis could explain in a natural way the construction of complex receptive fields from simple ones and of long oriented receptive fields from shorter ones.

When reconstructed from tangential sections, the HRP-labeled pattern forms a series of stripes. Humphrey *et al.* (2, 3) have shown that cells of a given orientation preference are arranged in a pattern of roughly parallel stripes, about 0.5 mm apart, in area 17 of the tree shrew. Rockland and Lund commented on the fact that the pattern they have observed closely resembles these orientation stripes, and they suggested that there might be a functional relationship between the two patterns. Let us consider what this might be.

Possible Explanations. The HRP pattern presumably reflects the connections of cells within the site of the HRP injection (we consider this in more detail later). Suppose a cell connects to all other cells nearby that have a similar orientation preference. Because the injection sites are considerably larger than the spacing between orientation columns and are likely to include cells of all orientations in roughly equal proportions, it is clear that, if all cells of similar orientation were interconnected, one would get a fairly uniform pattern of HRP, radiating away from the injection site. So this cannot account for the patchiness of the HRP pattern.

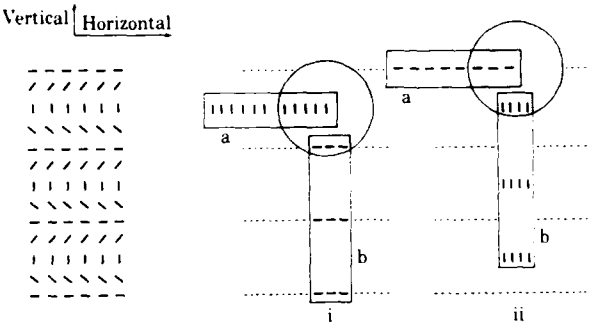


FIG. 1. An idealized set of orientation stripes is sketched on the left, the stripes following what corresponds to the horizontal direction in visual space. The stripes are assumed to continue running through the two adjacent sketches, with the horizontal stripe lightly indicated. Two separate cases are illustrated. In each the HRP injection is represented by a stippled circle. In the first case (*i*) we follow the convention that cells with similar orientations are joined within a narrow field whose direction is at right angles to the cells' orientation. Cells within the injection site having a vertical orientation are joined by a horizontal field (a) to others nearby, to give a continuous strip of label. Horizontally oriented cells connect within a vertical field (b) to patches of cells. The second case (*ii*) employs the other convention, that cells connect within a field of like orientation. We see that now it is the horizontally oriented cells that are joined to others within a continuous strip (a). The vertical cells connect to patches of cells (b).

At this point, it is reasonable to ask what purpose connections between cells with similar orientations might serve. Now, amongst cells in the striate cortex, one can broadly distinguish two types of receptive fields (4). One, the so-called simple type, gives a response to oriented bars at a particular location in space; the other, the complex type, responds to an oriented bar over a range of positions in space. Hubel and Wiesel (4) have suggested that complex fields could be made by joining together a number of simple fields side by side. To do this, one would need to join cells with like orientations. But instead of joining *all* such cells within some distance, one would generally need to join only those that lie within a strip running in a particular direction on the cortex.

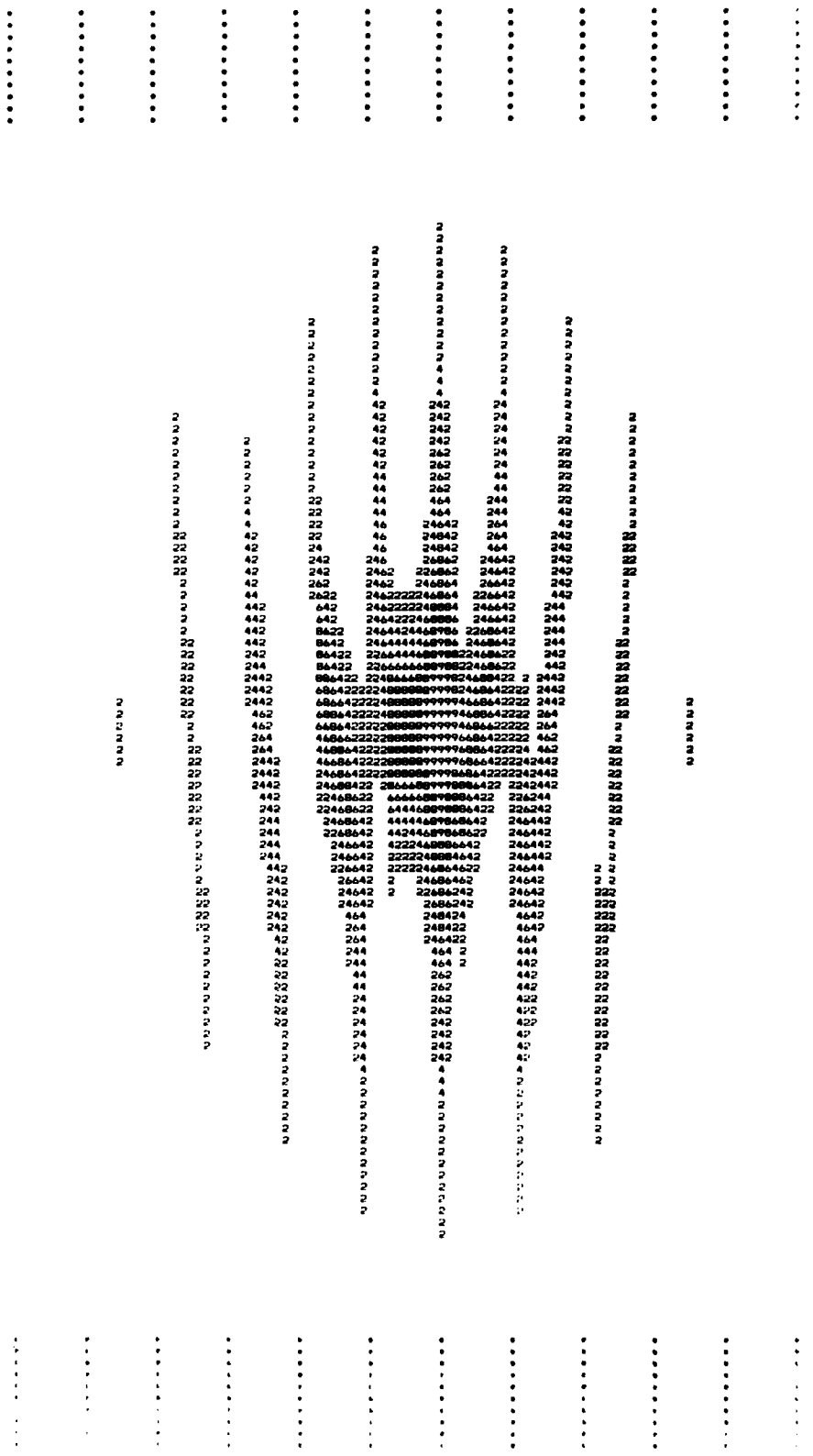
To see this, observe that at any point in the visual field a particular direction can be mapped onto the cortex as a direction in the tangential plane of the cortex. This is because, small distortions aside, the striate cortex is a map of the visual field. So, to construct a complex field by joining simple ones side by side,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviation: HRP, horseradish peroxidase.

the connections would have to be made in a direction at right angles to the cell's orientation preference, mapped onto the cortical surface.

Another possible role for the cortical connections we are considering would be in constructing very elongated fields by joining shorter ones end to end. A cell with such a field would then receive axon collaterals from distant cells having shorter receptive fields. But here the axons would have to follow a direction



roughly parallel to the cell's orientation preference, instead of at right angles, as for complex cells.

Our proposal is, then, that the HRP pattern is created by axons joining cells of like orientations in a particular direction on the cortex. The relationship between this direction and the cell's orientation preference will depend upon the type of field being constructed, complex fields giving one rule and elongated simple cells another.

The Stripes Illustrated. These two possible rules are illustrated in a highly idealized way in Fig. 1. In this figure the orientation stripes have been drawn to run horizontally. As one moves vertically in the figure, across the stripes, the orientations rotate steadily in a fixed direction. In the tree shrew the stripes are about 0.5 mm apart and, near the 17/18 border, are arranged roughly perpendicular to this border. (The 17/18 border is the representation of the vertical meridian in the retinotopic map.)

The first possible rule is illustrated in case *i* of Fig. 1. Consider for the moment only cells with a *vertical* orientation preference (field a). We join up only such cells that lie in a horizontal direction in the cortex from the HRP injection. These can be seen to lie along an orientation stripe, so all the cells in this stripe will be joined. Now consider only those cells within the HRP injection site that have a *horizontal* preference (field b). If these are connected only with other similar cells that lie in a vertical direction in the cortex from the HRP patch, they do

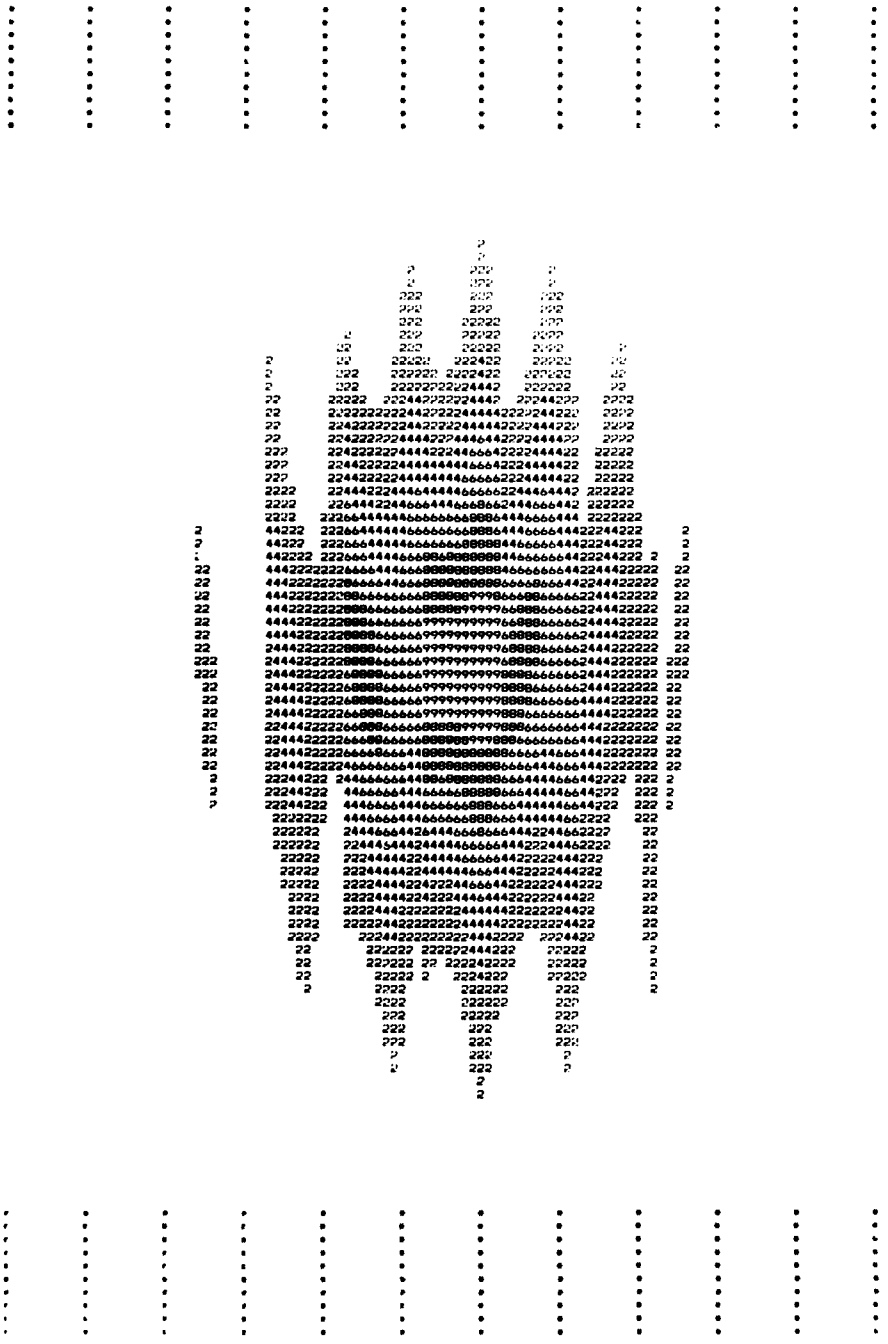


FIG. 2. Computer simulation of the stripes generated by using the rules illustrated in Fig. 1. Orientations were represented as parallel stripes on a regular grid, with a complete (180°) rotation in 10 steps on the grid. The orientation stripes run from the top to the bottom of the figure, as suggested by the short line segments above and below. If we adopt convention *i* of Fig. 1, then the lines drawn correspond to the vertical orientation; if we adopt convention *ii*, they are horizontal. We gave the connection field the shape of a Gaussian $G(u,v) = \exp[-(u^2/2a^2) - (v^2/2b^2)]$, in which the field is assumed to have its long axis in the u direction and its short axis in the v direction. The HRP injection site was represented by a 10 × 10 square of grid points, each point given a HRP concentration of c. Suppose the grid has coordinates (x, y), with the y coordinate parallel to the orientation stripes. If the point P with coordinates (x, y) in the grid had the same orientation as a point $Q_i = (x_i, y_i)$ in the injection site, then a HRP concentration $cG(x - x_i, y - y_i)$ was added to P. Such contributions were summed over all Q_i with the same orientation within the injection site. In these calculations we took $c = 1$, which meant that the HRP concentrations at each point lay in the range 0–10, the integral parts of which are shown by the digits in the printout. (Zero has been omitted and only the even numbers have been shown, except for 9.) (*Left*) The variance a was taken to be 20 steps, and b to be 5 steps. Because the repeat distance between stripes is 10 units, these correspond to distances of 2 column widths by 0.5 column width. (*Right*) Here we took $a = 15$, $b = 10$, corresponding to 1.5 by 1 column widths.

not form a continuous strip, because they cut across the orientation stripes. (We imagine that in such cases the axon collaterals run without interruption and ramify extensively only when they come to a group of cells that have the correct orientation preference. Only the regions where the axons are highly ramified will show up in the HRP pattern.)

If we now fill in the pattern by adding connections from cells of *all* orientations (always using the rule that the connections are made only in a direction of the cortex perpendicular to the orientation preference in question) it is easy to see that all these patches join up into a fairly continuous system of stripes.

The second rule is illustrated in case *ii* of Fig. 1. In this case the rule is that the straight connection lines are *parallel* to the orientation preference (mapped onto the cortex) of the cells that are joined together. By considering each orientation preference in turn we see that the HRP pattern will again be a pattern of stripes, though these stripes will interleave with the stripes formed in the first case considered. In both cases the stripes will be roughly parallel to the orientation stripes.

The general nature of the stripes can be seen rather more clearly in the computer simulation shown in Fig. 2. In this figure the idealized orientation stripes have been drawn vertically rather than horizontally as they were in Fig. 1. It has been computed with reasonable rules for the fall-off of connections with distance, which also allow for a certain scatter in the direction of connection. The connection rule is more diffusely oriented in Fig. 2 *Right* than it is in Fig. 2 *Left*; the details are given in the legend to the figure. It can be seen that in both patterns, but especially in Fig. 2 *Left*, the calculated stripes roughly follow the orientation stripes (which are marked only at the top and bottom of the figure to avoid confusion). They are usually not strictly parallel to them but are inclined at a small angle. Put in other words, the computed "HRP stripes" follow the orientation stripes but with a phase shift that depends on the angle joining a given point to the HRP injection site.

Complications. The orientation stripes in the tree shrew are not nearly as regular as the idealized grid we have used. They appear to consist of fairly large areas of nearly parallel stripes, interrupted by singularities [see figure 10 of Humphrey *et al.* (3)]. Moreover, during a long electrode penetration in a direction at right angles to the orientation stripes, the orientations encountered may steadily change in one direction up to a certain point, and then proceed to change in the opposite sense. Both singularities and changes in rotation would be expected to disturb the pattern. Singularities will introduce blind endings or branches, and HRP stripes will run together along a boundary where the rotation sense changes (Fig. 3). All one can say in general is that the HRP pattern will run roughly parallel to the orientation stripes, perhaps showing branches or anastomoses. Rockland and Lund's patterns fit this general description, having a similar spacing and direction to the orientation stripes, with some branches.

We have used here a very crude definition of connectivity. Rockland and Lund found that HRP was transported in an orthograde manner along the axons of cells whose bodies lay in the injection site. They also found a small number of retrogradely filled cell bodies. If HRP is carried to the body of such a cell along one axonal branch and is then carried further in the same direction along another axon branch, the label could move greater distances from the injection site. However, one might expect that only a small amount of label moves in this way.

Rocklund and Lund's pattern was found only in the superficial layers (II and III), the distribution being continuous in the deeper layers. The most likely candidates for the HRP-filled cells are small pyramids in layers II and III, whose axons travel distances of 1 mm or more in the upper layers. The size of the oriented connection field we have used is therefore reasonable, but our calculation does not allow for retrograde movement to the cell body and beyond, which, if appreciable, would spread the stripes yet wider.

The Cat and the Tree Shrew. To interpret the system of connections for the tree shrew, we consider first what is known for another animal, the cat. As Gilbert (5) has shown, the cells in layers II and III in area 17 of the cat have receptive field widths very similar to those found in layer IV, which is the layer that receives most of the input from the lateral geniculate nucleus. The principal difference between cells in layers II and III and those in layer IV is that many of the former have complex receptive fields, whereas the latter are mostly simple. Gilbert (5) found that the length of bars that provided the optimal stimulus for complex cells in layers II and III was about the same as that for the simple cells in layer IV. He therefore suggested, following a proposal by Hubel and Wiesel (4), that a side-by-side joining-up operation is the main transformation that occurs in going from layer IV up to the supragranular layers. If the pattern of intracortcal wiring in the tree shrew is similar to that in the cat (which there is every reason to believe is the case), we propose that the axon system shown up by the HRP pattern in the tree shrew carries out a similar operation in the upper layers of the cortex, and it consists of axons running at right angles to the preferred orientations of both their cells of origin and of the cells with which they connect.

The cat appears to have a system of orientation stripes like that in the tree shrew (6), so we might expect to find a pattern like Rockland and Lund's in the upper layers of the cat's striate cortex. However, we might also look for another type of connectivity. Gilbert and Wiesel (7) recorded from cells in area 17 of the cat, and then filled them with HRP. They described a long axon from a layer V cell that traveled a great distance through layer VI, and they noticed that the orientation of the axon in the cortex paralleled the preferred orientation of its parent cell. They conjectured that this arrangement could be involved in making the very long receptive fields that are characteristic of many layer VI cells. So, instead of joining oriented fields side-by-side to make complex fields, we should also consider that short fields may be strung end-to-end. In this case, we should expect to find that a HRP injection gave a stripe-like

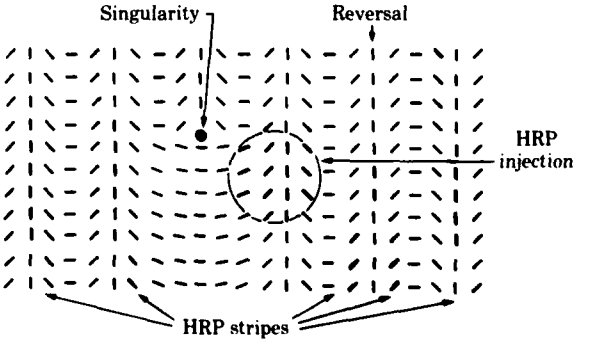


FIG. 3. The types of HRP pattern expected, using convention *i* of Fig. 1, when the orientation pattern is a little more complicated than the idealized one used before. In this figure the orientation stripes run from top to bottom. The HRP injection site is represented by the shaded circle. The HRP stripes (also shaded) were estimated, not calculated. On the left a singularity, S, is shown, of the kind postulated in figure 10 of Humphrey *et al.* (3). On the right there is a simple reversal so that the rate of change of orientation, going from left to right, changes sign at this point. As can be seen, the pattern of HRP stripes is distorted in these cases. Other cases, such as inclined reversals, give even more complicated distortions. Nevertheless, the general character and trend of the orientation stripes and the HRP stripes are similar.

pattern, but with the stripes shifted through one half-period of the underlying orientation pattern, because the cortical connections are turned through a right angle (Fig. 1, case *ii*, a and b). This pattern might be seen in the lower layers of the cat's striate cortex.

We should not expect to find a pattern of this kind in the lower layers of the tree shrew cortex, however, because it appears to lack cells with the very elongated receptive fields of the type found in layer VI of the cat. In fact, cells in the lower layers of the tree shrew's area 17 are less oriented than those in the supragranular layers (2). This may explain why Rockland and Lund's pattern shows no patchiness in layers V and VI. The computed HRP pattern for cells with less elongated fields has less conspicuous stripes (Fig. 2 *Right*).

Other Possibilities. We have now considered two ways in which systems of oriented axons might arise in the striate cortex. Another possibility is that there are axons leading to or from inhibitory cells, creating either the inhibitory flanks of receptive fields or the end-stopping that, in the cat, is common in layer IV and above (7). The former would yield a pattern similar to the one proposed for making complex fields, the latter would create the 90° phase-shifted version. We have thus considered four ways, excitatory and inhibitory, in which the appropriate kinds of connections might arise. Which of these, if any, is involved in generating the HRP pattern remains to be seen. But it should perhaps be pointed out that if two such mechanisms operate at once, their combined effect might either enhance or conceal the HRP stripes, depending whether they are in phase or 90° out of phase.

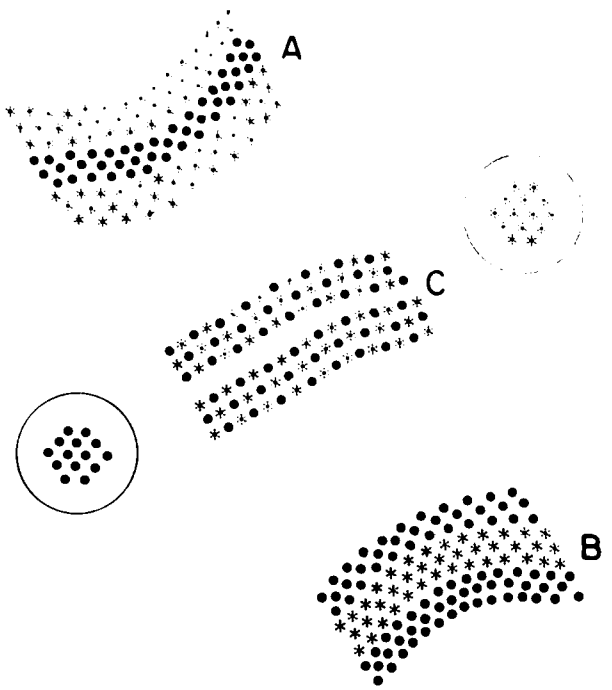


FIG. 4. Two injection sites, each injected with a different tracer, are represented by the two large circles. For simplicity the expected HRP stripes are shown only in the regions labeled A, B, and C. To distinguish the contributions from each site to neighboring cells, we label one with small circles, the other with asterisks. At a point C on the line joining the two sites, the connection field joining a cell to one of the sites has the same orientation as the field joining that cell to the other site. So any cell that is labeled by one site will be labeled (even if only weakly) by the other site, and the stripes of label will superpose (circles and asterisks intermingled). At either A or B, a cell that connects to one site cannot connect to the other, because the two connection fields would be at 90° to each other. Instead, each site will label interleaving stripes that differ by approximately 90° in orientation (circles and asterisks filling in adjacent bands). This argument holds no matter where the two injections are made, provided the retinotopic map is not grossly distorted on the large scale.

Conclusions. If our general hypothesis is correct, then axons in a tangential section, prepared as in Rockland and Lund's experiments, should be oriented and be connected to other cells whose axons have the same orientation preference. If axons are involved in creating complex fields, or flank inhibition, the direction of a cell's axon should correspond to an orientation in visual space that is at right angles to the preferred orientation of the cell. If axons are used to generate elongated fields, or end-stopping, then the axons' direction should coincide with the preferred orientation of the cells.

In either case, the HRP stripes should run roughly parallel to the orientation stripes, with perturbations near to singularities, or boundaries where there is a reversal of the direction of change of orientation (Fig. 3). There should be a variable phase shift between orientation and HRP stripes, depending upon the direction from a given point to the injection site. To illustrate this, suppose two injections of different tracer molecules (HRP and some other tracer transported in a similar manner) were made into the striate cortex, we should expect that on a line joining the two injection sites (e.g., region C in Fig. 4), stripes of label from the two sites would superpose. But near those points at which the sites subtend a right angle, (regions A and B in Fig. 4), the stripes should be 90° out of phase and interleaved.

In summary, there appear to be three main experimental questions: (*i*) Is the pattern found by Rockland and Lund due to stripes of cells with long collaterals interleaved with stripes containing no short collaterals? Or, as we suggest here, do all regions of area 17 have long collaterals but only in particular directions from any one small patch of cortex? (*ii*) If the latter is correct, what is the relationship between the orientation preference of a cell and the orientation of its collaterals? (*iii*) What is the functional significance of these long connections?

Naturally the answers to these questions may differ from species to species and from one cortical layer to another. However, in all cases it seems likely that the questions can be answered with present experimental methods.

We thank Drs. Kathleen Rockland and Jennifer Lund for sending us their paper in advance of publication. We also thank them and Dr. Maxwell Cowan for many helpful comments on various drafts of this paper. This work was supported by the Samuel Roberts Noble Foundation, the J. W. Kieckhefer Foundation, the Eugene and Estelle Ferkhauf Foundation, and U.S. Air Force Grant AF OSR-82-0042.

1. Rockland, K. S. & Lund, J. S. (1982) *Science* **215**, 1532–1534.
2. Humphrey, A. L. & Norton, T. T. (1980) *J. Comp. Neurol.* **192**, 531–547.
3. Humphrey, A. L., Skeen, L. C. & Norton, T. T. (1980) *J. Comp. Neurol.* **192**, 549–566.
4. Hubel, D. H. & Wiesel, T. N. (1962) *J. Physiol. (London)* **160**, 106–154.
5. Gilbert, C. D. (1977) *J. Physiol. (London)* **268**, 291–421.
6. Singer, W., Freeman, B. & Rauschecker, J. (1981) *Exp. Brain Res.* **41**, 199–215.
7. Gilbert, C. D. & Wiesel, T. N. (1979) *Nature (London)* **280**, 120–125.

November 23, 1982
Submitted to Nature

ON THE FUNCTION OF DREAM SLEEP

Francis Crick and Graeme Mitchison*
The Salk Institute
10010 N. Torrey Pines Road
La Jolla, California 92037

* Present address: MRC Laboratory of Molecular Biology and the Kenneth Craik Laboratory, Cambridge, England.

SUMMARY

We propose that the function of dream sleep (more properly called REM sleep) is to maintain a delicate balance between excitation and inhibition in the cerebral cortex. We postulate that this is done in REM sleep by a reverse learning mechanism, so that the trace in the brain of the unconscious dream is weakened, rather than strengthened, by the dream.

Mankind has always been fascinated by dreams. As might be expected, there have been many attempts to assign a purpose or significance to them. Although we dream for one or two hours every night, we are not consciously aware of most of our dreams. Earlier thinkers, such as Freud, did not know this. Modern theories (not reviewed here in detail) have usually proposed that sleep and dreams save energy or have various restorative functions, either to replenish the brain biochemically in some way, or to reclassify or reorder the information stored in it. Such processes may be postulated to occur during brain development as well as during adult life.

Sleep is of several kinds. Dream sleep, or rapid eye movement (REM) sleep, is predominantly found in placental mammals and seems to be associated with homeothermy and the possession of an appreciable neocortex. It is not unimportant because of the appreciable amount of time we spend in this peculiar state.

We propose here a new explanation for the function of REM sleep. The basis of our theory is the assumption that in placental mammals the cortical system (the cerebral cortex and some of its associated subcortical structures) is normally in a delicate state of balance between excitation and inhibition. We believe this balance is likely to be upset both by the growth of the brain and also by some of the modifications produced in it by experience. For this reason a special mechanism is needed to keep it balanced. We propose that this balancing is done during REM sleep by an active process which is, loosely speaking, the opposite of learning. We call this "reverse learning" or "unlearning". This mechanism, which is not the same as normal forgetting or some non-active decay process, is explained in more detail below. Without it we believe that the mammalian cortex would either become unstable or could not perform so well.

We first describe our ideas about the cortex. Next we outline what is known about REM sleep. (For general accounts, see the first two references.) We then describe our postulated mechanism and how it might be tested. Finally we discuss various implications of our ideas.

The General Nature of the Cortex

The cortex consists of two separate sheets of neural tissue, one on each side of the head. The neocortex, which has a characteristic layered structure, is found only in mammals (for a recent survey, see reference 3). If allowance is made for body weight, it is larger in primates than in most other mammals and larger in man than in other primates. It makes up a substantial fraction of the human brain.

Different areas of the cortex perform different functions, some being mainly associated with vision, touch, etc., while

others appear to process more complex information. The exact function of the neocortex is unknown but it appears to be closely associated with higher mental activities. It seems likely that it has evolved to perform in a rather special way.

In examining the neuroanatomy of the neocortex one is struck by the very large number of local axons and axon collaterals. In any area of the cortex the great majority of synapses come from axons originating locally and running within it. Those synapses due to axons which enter a part of the cortex from outside form a minority[4].

Synapses can usually be classified as one or another of two main types, known as Type I and Type II. It is believed that Type I synapses excite and Type II inhibit. Since there is so much local feedback in the cortex, one would expect that the inhibition would be large enough to balance the excitation, or the system would become unstable and perhaps go into oscillation. Thus at first sight one might guess that the two types of synapse would occur in approximately equal numbers. This is not the case. The Type II (inhibitory) synapses are in a minority. However most of them are placed in more effective positions than the Type I's. For pyramidal cells, which make up about two-thirds of the total, the Type II synapses are located mainly on the more proximal parts of the dendrites, on the soma and on the axon initial segment. For this reason the Type II synapses, though fewer in number, can probably balance the Type I synapses.

In various conditions, such as epilepsy, migraine and certain kinds of drug-induced hallucination[5] parts of the cortex appear to go into large-amplitude instabilities[6]. This is not surprising, considering the extensive collateral excitation just described and the many reciprocal connections with the thalamus, the main organ of entry to the cortex.

Our basic hypothesis is that when the cortex is operating normally, its many parameters are tuned so that it is not far from instability in order in some way to improve its function. For example, it is possible that a more finely-tuned system can build up the neural representation of a percept more quickly than a less finely-tuned one or that it needs less neocortex to achieve the same performance. The evidence in the previous paragraph clearly shows that the cortical system can become unstable but does not establish how close it normally is to instability.

If the cortex were hard-wired during embryogenesis to an exactly predetermined pattern of synaptic connections, the burden of balancing the cortex would have to be undertaken by the genes alone. Although there is considerable evidence for specificity in the cortical wiring, it is likely that much of the details of the synaptic connections -- their exact locations and their

strength -- are made in a semi-random manner and refined by experience. This is almost a necessity in an organism which is capable of learning very large amounts of novel information. Thus it seems likely that both during cortical growth and also in facing the experiences of adult life the cortex would have difficulty in finely tuning itself. This is because the system is so complex and so inter-connected that to tune it up in one sub-mode of operation is likely to make it unstable in another, related one.

How would one attempt to balance such a system? We suggest the following. The major inputs and outputs of the system should be turned off, so that the system is largely isolated. It should then be given successive "random" activations, from internal sources, so that any incipient parasitic modes would be excited, especially if the general balance of excitation to inhibition had been temporarily tilted towards excitation. Some mechanism is then needed to make changes so that these potentially parasitic modes are damped down. Such a rough outline description immediately reminds one of REM sleep and the hallucinoid dreams associated with it.

The Nature of REM Sleep

It was discovered in the fifties that in mammals there are two main types of sleep. Periods of REM sleep ("REM" for Rapid Eye Movement -- also called D sleep or paradoxical sleep) alternate with periods of non-REM sleep (also called S sleep, slow wave sleep, or orthodox sleep) of which four stages of increasing depth of sleep are usually distinguished. During REM periods the sleeping animal may be more difficult to arouse and many of its muscles, especially its head and neck muscles, are more relaxed than in non-REM sleep. Its cortex, as judged by the electroencephalogram (EEG) and by the rapid movement of his eyes beneath closed lids, appears to be very active. On the other hand, the monoamine neurons in the brain stem reduce their firing rates in REM sleep to only a few percent of the corresponding rate in the waking state[7].

Another major difference between REM and non-REM sleep lies in the dreams associated with them. For most people the few dreams found in non-REM sleep tend to have a rather thought-like character. During REM sleep, on the other hand, dreams occur more frequently and usually have a perceptual vividness and the illogical episodic character with which we are all familiar. A human adult usually spends a total of 1-1/2 to 2 hours each night in REM sleep, spread over several periods. The evidence suggests that most of the dreams during these REM periods do not reach normal consciousness, dreams being remembered only if the sleeper awakes while dreaming. Even then the memory of a dream is usually very transient, fading quickly if no effort is made to remember it by rehearsing its content.

A most remarkable finding is that newly-born humans may have as much as 8 hours of REM sleep per day[8]. There is also evidence to suggest that in the womb, especially in the third trimester, REM sleep occurs even more frequently. This large amount of REM sleep before and after birth is also found in other mammals. These facts were unknown to Freud.

All placental mammals examined, including primitive marsupials such as the opposum, show periods of REM sleep[9,10]. Even an animal like a mole, which can hardly move its eyes, shows the characteristic EEG of REM sleep. Birds have a very small amount of REM, occupying perhaps 5% of their sleep[11]. There are no very convincing reports of REM sleep (as judged by the EEG) in reptiles, amphibia or fish.

If an animal is deprived of REM sleep for one or more nights (but allowed non-REM sleep) then it will have proportionately more REM sleep in subsequent nights[12,13].

All this evidence suggests that REM sleep has an important function, at least for mammals. Since the majority of dreams are not remembered, that function is more likely to be associated with the unconscious dreaming process -- that is, with REM sleep without awakening -- rather than with the few dreams which are recalled.

It has been shown that during REM sleep the forebrain is periodically and widely stimulated by the brain stem. This activity in the brain stem can happen even in the absence of the cortex. Hobson and McCarley[14], following the pioneer work of Jouvet[15], have postulated a Dream State Generator which lies mainly in the pontine reticular formation, an important component being the giant cells of the pontine tegmentum. They propose that the activity of such cells is the cause of both rapid eye movements and the periodic intrusion of new subject matter into our hallucinoid dreams. Our proposals are based on this idea.

In summary, the evidence suggests that in REM sleep the brain is especially isolated from its normal input and output channels and that it is very active, this activity being promoted by rather non-specific signals from the brain stem and reflected in the unconscious equivalent of dreaming, which only becomes conscious if the sleeper awakes.

The postulated mechanism

We need a mechanism which will balance the cortical system, especially when this balance has been disturbed either by growth of the brain (when new connections are constantly being made) or by the modifications produced by experience. The mechanism we propose is based on the more or less random

stimulation of the forebrain by the brain stem, described above. We postulate that this will tend to stimulate modes of brain activity which are too prone to be set off by random noise rather than by highly structured specific signals. We further postulate a reverse learning mechanism which will modify the cortex (for example, by altering the strengths of individual synapses) in such a way that this particular activity is less likely in the future. For example, if a synapse needs to be strengthened in order to remember something, then in reverse learning it would be weakened. Put more loosely, we suggest that in REM sleep we unlearn our unconscious dreams. "We dream in order to forget."

If there is indeed a mechanism for reverse learning, many questions arise about its character. Does it act upon the same mechanism as normal learning (whatever that is) or is a special, quite separate, mechanism involved? If so, what is its neural correlate? The hippocampus has been implicated in learning, and in accessing certain forms of longer term memory. Is it also involved in reverse learning? At this stage we do not wish to make any suggestion on these points.

In its simplest form our theory postulates there is no intelligent supervisor inside the brain which decides in detail which potential neural activities should be left untouched and which should be damped down. This choice is made solely by the response of the forebrain to the relatively non-specific signals from the brain stem. In very general terms, the brain stem gives the forebrain a varied pattern of bangs. Any resulting activity is then modified so that it is less likely to occur in the future.

It would of course be possible to postulate a more complex mechanism. For example, in REM sleep, especially in early development, there could be innate testing programs, together with a "supervisor" to decide what to store and what to erase, depending on the result of the tests. Various workers have made proposals along these lines[16-21]. As far as we know, nobody has previously suggested that the testing procedure involves the removal of potential instabilities.

It has been customary to believe that during an unconscious dream the content of the dream is stored in some form of very short-term "memory" but that the mechanism for transferring it into longer term memory is inoperative. We normally become conscious of our dreams only if we wake up while dreaming is in progress. If we then pay attention to our dream, some of its content can be maintained in very short-term memory and may eventually be transferred to longer term memory as the transfer mechanism becomes activated. Otherwise our dream fades. Thus we can speak of forgetting our dreams, meaning we know that we had a dream, but are somewhat uncertain of its content.

This forgetting of a dream, which has often been remarked on, does not necessarily involve our postulated reverse learning process. The latter is a positive mechanism which does not merely fail to alter synaptic strengths (or other long-lasting brain parameters) but changes them so that the dream is not just forgotten but actively "unlearned". The result is that the dream, or some of the elements of it, are less likely to recur in the future.

The terms "reverse learning" or "unlearning" are not ideal because they rather imply that one has to learn something first in order to unlearn it. What does a fetus "learn" that has to be unlearned? Our answer is that, during development, the semi-random process of making synaptic connections is likely to produce instabilities. It is these instabilities which must be "unlearned" in order to obtain a stable, finely-tuned system.

We need some explanation for recurrent dreams. We propose the ad hoc hypothesis that a recurrent dream is one which, for one reason or another, tends to wake up the sleeper, perhaps because of the anxiety often associated with them. This will have the effect that the learning process changes sign, passing from reverse learning to positive learning, so that the underlying instability remains, and so a similar dream is likely to occur on some later occasion. This mechanism does postulate a supervisor of a kind but its sole function is to decide whether the sleeper should wake up or not. Thus for a dream to become recurrent it must have two properties. It must be related to a potentially parasitic mode and it must wake up the dreamer in such a way that he remembers it rather vividly.

Our theory, in its present state, says nothing about the function of non-REM sleep. These stages of sleep usually have less of the hallucinoid type of dream which we associate with our reverse learning mechanism. Non-REM sleep is likely to have the restorative function often postulated for it but it may also have some informational function. For example, it might be used for the process of "consolidating" memory in some way. It is worth noting that the first REM period of the night is normally preceded by a substantial period of non-REM sleep.

Testing the Theory

As far as we can tell, our theory is compatible with a large amount of experimental data. Starting from a plausible hypothesis about cortical function, it explains in an effortless way both the need of REM sleep in adult life and the large amount of it during the development of the brain. We believe no previous theory explains this distribution of REM sleep in such a simple manner. Any purely psychological theory is hard pressed to explain the large amount of REM sleep in the womb, and any purely developmental theory must account for the quite

appreciable amount of REM sleep in adult life. Our theory accounts for both. It is also compatible with the hallucinoid nature of REM dreams.

The effects of REM sleep deprivation are harder to explain. It is well established that REM deprivation often produces a rebound -- more REM sleep than usual occurs when the subject is eventually allowed to sleep without interruption. We would have expected that REM deprivation, if severe enough, might cause hallucinations -- that is, structured visual and auditory responses to "noise" -- and perhaps delusions and obsessions. There is a little evidence for this[22], but usually the effects are either small or absent[23]. This is partly because it is extremely difficult to produce long periods of complete REM deprivation in humans by selective arousal. After a week or two it becomes almost impossible to awaken them promptly at every onset of REM sleep, so that prolonged experiments have not been done. One cannot help but wonder whether similar experiments on food deprivation might lead to the conclusion (if unsupported by other evidence) that food also had no essential function. However REM deprivation in animals does appear to lower the threshold for cortical instability produced by electroconvulsive shock[24-27], which is what we might expect. REM deprivation in humans sometimes produces irritability and an inability to concentrate. One might suggest that these are the effects of the attention mechanism being forced to subdue sub-threshold parasitic modes which would otherwise break into consciousness. REM deprivation can also allow feelings and wishes to appear which had previously been kept out of consciousness[28], or, in certain subjects, can show changes towards increased internal fantasy during waking[29].

A further difficulty is that some drugs, such as certain monoamine oxidase inhibitors, appear to prevent REM sleep entirely[30] (or at least the external signs of it) without producing very obvious psychological deficits. This is a difficulty for <u>any</u> theory which assumes REM sleep is important and runs in the face of all the other evidence about it. We can only plead that such drugs may have complicated side effects which make the observations misleading, or that the stabilizing effect of the drug compensates for the lost stabilization achieved in REM.

A <u>direct</u> test of our postulated reverse learning mechanism seems extremely difficult. It would be necessary to show that our unconscious dreams* (dreams we do not remember) reduce the probability of such

---

*A new word is really needed for this concept. With some hesitation we suggest "remination".

thoughts occuring in the future. This is far beyond the methods we have available today. It would be interesting to know if the threshold for hallucination, induced by drugs or other means, is lowered as the result of REM deprivation. Another approach would be to look for the structural and chemical correlates of the postulated reverse learning mechanism, but exactly how to do this is at the moment unclear. Without further evidence of this kind our theory must be regarded as speculative.

Some useful insights might come from neural modelling. This approach has its limitations, since it is difficult to produce realistic models and even more difficult to simulate them effectively, especially if the hypothetical neural nets approach a realistic size, when the computational time becomes prohibitively long. However such theoretical studies should at least reveal some of the types of networks which would benefit from our proposed mechanism. They might also help to give more life to our otherwise rather vague characterization of the cortical system. Put another way, our postulate that the cortex is finely balanced, if made more precise, might help us to understand better some of its neurophysiology and neuroanatomy.

Another approach would be comparative studies. There is one mammal which, though possessing a well-developed neocortex[31], appears not to show any signs of REM sleep (at least in young adults), even though it exhibits normal non-REM sleep[32]. This is the echidna Tachyglossus aculeatus (the spiny anteater) found in Australia. The Echidnas and the duck-billed platypus are primitive egg-laying mammals (monotremes).

Tachyglossus can be studied in captivity. It might be rewarding to examine in more detail its behavior, neurophysiology and neuroanatomy compared to a primitive placental mammal, such as a hedgehog, which does show REM sleep. If REM sleep serves an important function, this should be reflected in some way in its absence in the spiny anteater.

Possible Implications

If it turns out that our ideas are broadly correct, they could help us to understand the evolution of the neocortex which is so typical of mammals. It seems likely that in order for a highly tuned system to evolve at least two requirements are necessary: a fairly constant internal temperature, so that its balance is not disturbed by temperature fluctuations, and in addition a cleaning-up mechanism, to preserve this balance and to remove potentially parasitic modes. In short, without REM dreams the cortical system could not have evolved into the highly refined instrument we have today.

If the reverse learning mechanism we have postulated exists, one might wonder what effects its failure might have. A

complete failure might lead to such grave disturbances -- a state of almost perpetual instability -- that it would probably be severely selected against. A partial failure should produce unwanted responses to random noise, perhaps as hallucinations, delusions, and obsessions, and produce a state not unlike certain forms of schizophrenia. It has been postulated before that there might be a relation between REM sleep and schizophrenia, but studies have shown that there is little or no connection between the outward signs of REM sleep and schizophrenia[33]. However a partial failure of the reverse learning mechanism would not necessarily alter the amount of REM sleep, since the control mechanisms for the occurrence of REM sleep might be somewhat distinct from the reverse learning process itself. Thus the possibility that some forms of schizophrenia might be caused by a defect in the reverse learning process should not be overlooked.

If our ideas are correct, attempting to remember one's dreams should normally be discouraged, because such remembering merely helps to retain patterns of thought which are better forgotten, since these are the very patterns the organism was attempting to damp down.

Finally we should remark that even if it turns out that our ideas are wrong and that Nature does not employ the reverse learning mechanism we have postulated, the process may well be useful for artificial intelligence machines of the future, especially those having extensive parallel processing, a learning mechanism and a certain amount of randomness in their construction.

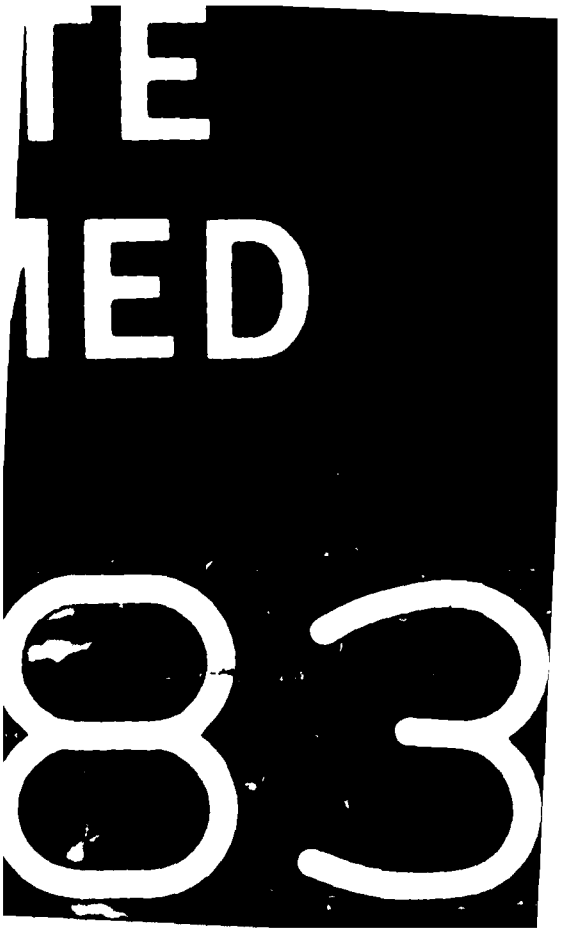
Acknowledgements

We thank our resident and visiting colleagues at The Salk Institute for many useful discussions. We are especially grateful to Dr. Allan Hobson and Dr. Jim Horne who made detailed comments on our draft manuscript. This work has been supported by the J. W. Kieckhefer Foundation, Samuel Roberts Noble Foundation, U.S. Air Force Grant number AFOSR-82-0042 and the System Development Foundation.

References

1. Hartmann, E. L. The Functions of Sleep (Yale Univ Press, London, 1973).
2. Cartwright, R. D. A Primer on Sleep and Dreaming (Addison-Wesley Publishing Co., Reading, Mass., 1978).
3. Schmitt, F. O., Worden, F. G., Adelman, G. and Dennis, S. G. (Eds.) The Organization of the Cerebral Cortex (The MIT Press, Cambridge, Mass., 1981).
4. Braitenberg, V. in Architectonics of the Cerebral Cortex (Eds. Brazier, M. A. B. and Petsche, H.) 443-465 (Raven Press, New York, 1978).
5. Siegel, R. K. Sci Amer 237, 132-141 (1977).
6. Ermentrout, G. B. and Cowan, J. D. Biol Cybernetics 34, 137-150 (1979).
7. Hobson, J. A., McCarley, R. W. and Nelson, J. P. (in the press).
8. Roffwarg, H. P., Muzio, J. N. and Dement, W. C. Science 152, 604-619 (1966).
9. Allison, T. and Van Twyver, H. Natural History 79, 56-65 (1970).
10. Allison, T. and Cicchetti, D. V. Science 194, 732-734 (1976).
11. Klein, M., Michel, F. and Jouvet, M. C R Soc Biol 158, 99-103 (1964).
12. Dement, W. Science 131, 1705-1707 (1960).
13. Kales, A., Hoedemacher, F., Jacobson, A. and Lichtenstein, E. Nature 204, 1337-1338 (1964).
14. Jouvet, M. Arch Ital Biol 100, 125-206 (1962).
15. Hobson, J. A. and McCarley, R. W. Am J Psychiatry 134, 1335-1348 (1977).
16. Newman, E. A. and Evans, C. R. Nature 206, 534 (1965).
17. Gaarder, K. Arch Gen Psychiat 14, 253-260 (1966).
18. Dewan, E. M. in Sleep and Dreaming (Ed. Hartmann, E.) 295-307 (Little-Brown and Co., Boston, 1970).
19. Greenberg, R. and Pearlman, C. Persp Biol Med Vol 7 No 4, 513-521 (1974).
20. Jouvet, M. in Progress in Brain Research, Vol 53 (Eds. McConnell, P. S., Boer, G. J., Romijn, H. J., van de Pol, N. E. and Corner, M. A.) 331-346 (Elsevier/North-Holland Biomedical Press, Amsterdam-New York, 1980).
21. Hobson, J. A. in Brain Mechanisms and Perceptual Awareness (Ed. Pompeiano, O. and Marsan, C. A.) 379-404 (Raven Press, New York, 1981).
22. Dement, W. in Academy of Psychoanalysis: Science and Psychoanalysis (Ed. Masserman, J.) 7, 129-184 (Grune & Stratton, Inc., New York, 1964).
23. Vogel, G. W. Arch Gen Psychiat 18, 312-329 (1968).
24. Cohen, H. B. and Dement, W. C. Science 150, 1318-1319 (1965).
25. Hartmann, E., Marcus, J. and Leinoff, A. Psychonom Sci 13, 141-142 (1968).
26. Cohen, H., Thomas, J. and Dement, W. C. Brain Research 19,

317-321 (1970).
27. Cohen, H. B. and Dement, W. C. Brain Research 22, 421-422 (1970).
28. Greenberg, R., Pearlman, C., Fingar, R., Kantrowitz, J. and Kawliche, S. Br J Med Psychol 43, 1-11 (1970).
29. Cartwright, R. D. and Ratzel, R. W. Arch Gen Psychiat 27, 277-280 (1972).
30. Wyatt, R. J., Kupfer, D. J., Scott, J., Robinson, D. S. and Snyder, F. Psychopharmacologia 15, 236-244 (1969).
31. Griffiths, M. The Biology of Monotremes (Academic Press, New York, 1978).
32. Allison, T., Van Twyver, H. and Goff, W. R. Arch Ital Biol 110, 145-184 (1972).
33. Vogel, G. W. and Traub, A. C. Arch Gen Psychiat 18, 287-300 (1968).


TE
MED
83